



# Digital Video Library

Informedia-II: Integrated Video Information  
Extraction and Synthesis for Adaptive  
Presentation and Summarization from  
Distributed Libraries

Howard D. Wactlar  
Carnegie Mellon University



# Digital Video Library

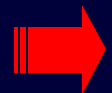
## Informedia-II: Auto-Summarization and Visualization Over Multiple Video Documents and Libraries

Howard D. Wactlar  
Carnegie Mellon University

# Why Video ?

---

- Our society -
  - *captures* its experience,
  - *records* its accomplishments,
  - *portrays* its past
  - *informs* its masses
  - .....in audio and video
- For many, CNN has become the “publication of record”
- Video learning leverages “multiple intelligences” Gardner, 1993
- Video libraries will become an essential component of
  - formal, informal, and professional learning
  - distance education, telemedicine

 ***We need to provide for video at least the same capability for indexing, search and retrieval that will exist for text***

# Informedia-I Overview

---

## *Enable Search and Discovery in the Video Medium*

- Automated information and metadata extraction from video
- Full-content search and retrieval from *spoken language and visual documents*
- Integration of speech, image and natural language understanding for library creation and exploration
- Validation through user testbeds

 ***Establishment of large video libraries as a searchable information resource***

# Application of Diverse Technologies

- Speech understanding for automatically derived transcripts
- Image understanding for video “paragraphing”; face, text and object recognition
- Natural language for query expansion, topic detection and content summarization
- Human computer interaction for video display, navigation and reuse
  - 
  - 
  -
- ***Integration overcomes limitation of each***

# Key Challenges for Informedia-II

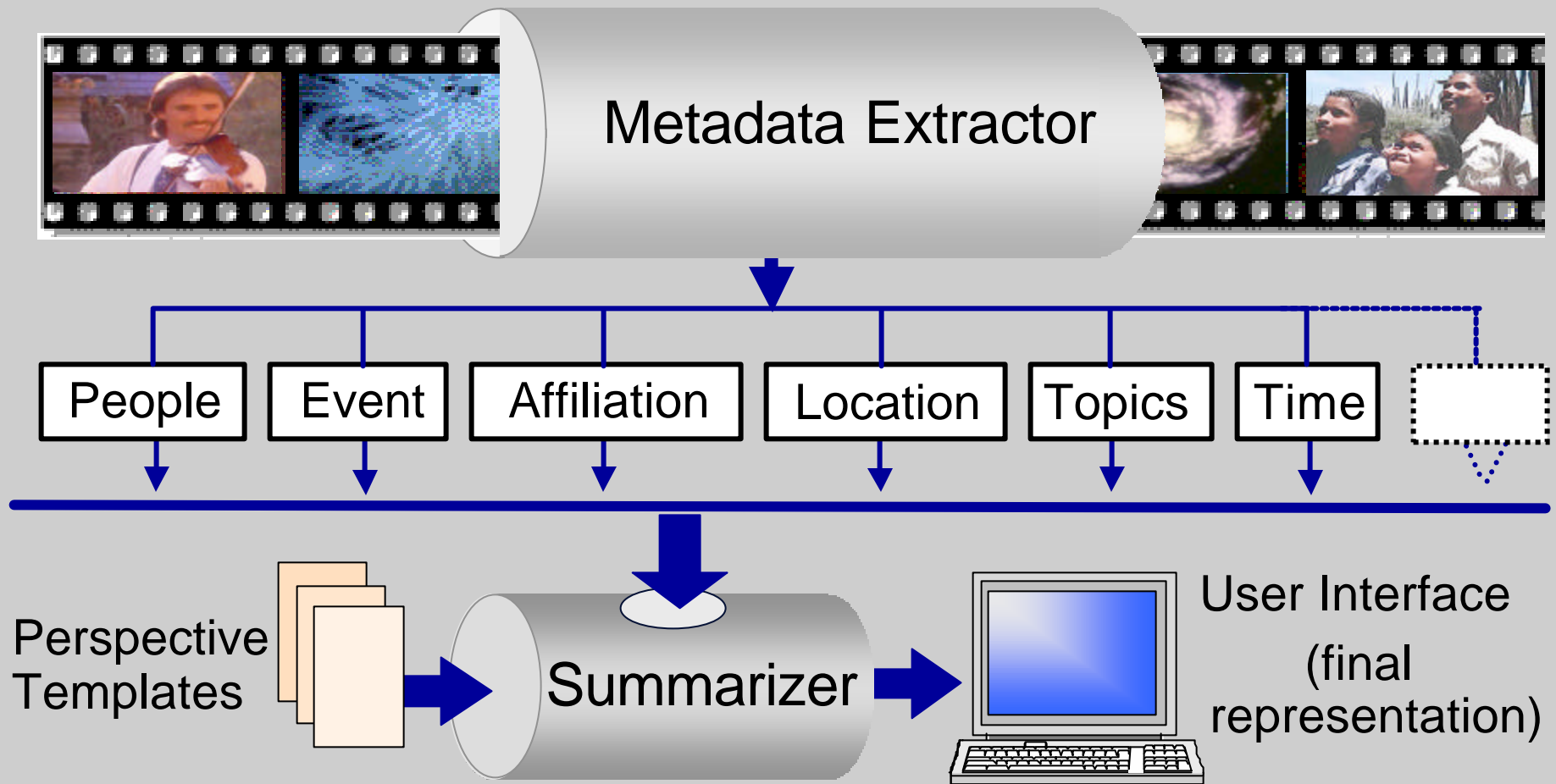
---

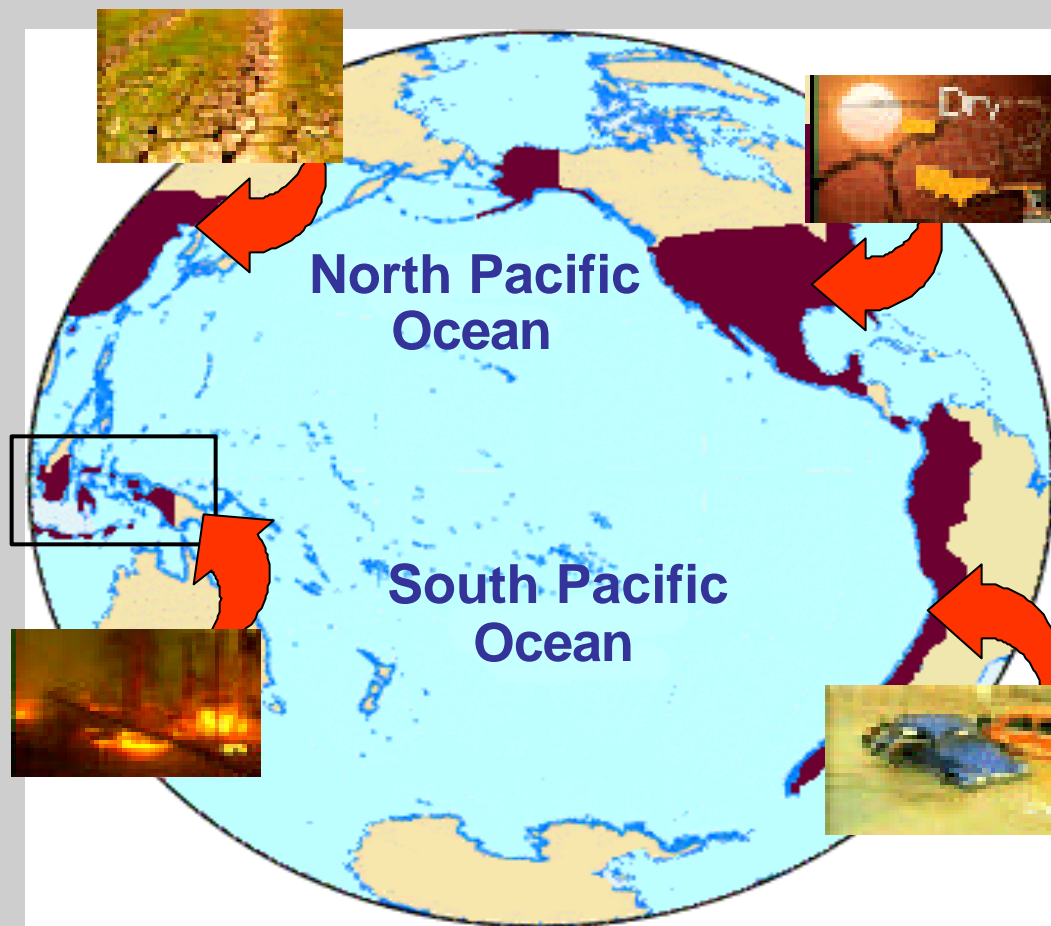
*Goal:* Auto-summarization and visualization of multiple audio & video documents, stories and reports

- create *collages*: by time, location, personalities, events
- adapted to *context of query*
- eliminating redundancy
- at any level of detail
- creating personality and event profiles on-demand

*Requisite subgoals:*

- establishing *geographic and temporal context*
  - geocoding all data by source and internal reference
  - extracting explicit and indirect time/date references
- video characterization and classification of image/verbal content
- integrating/scaling across *heterogeneous distributed collections*
  - establishing standards to enable combining remote content





Map collage emphasizing distribution by nation of  
“El Niño effects” with overlaid thumbnails





Map collage produced when user zooms into the boxed area around Indonesia.



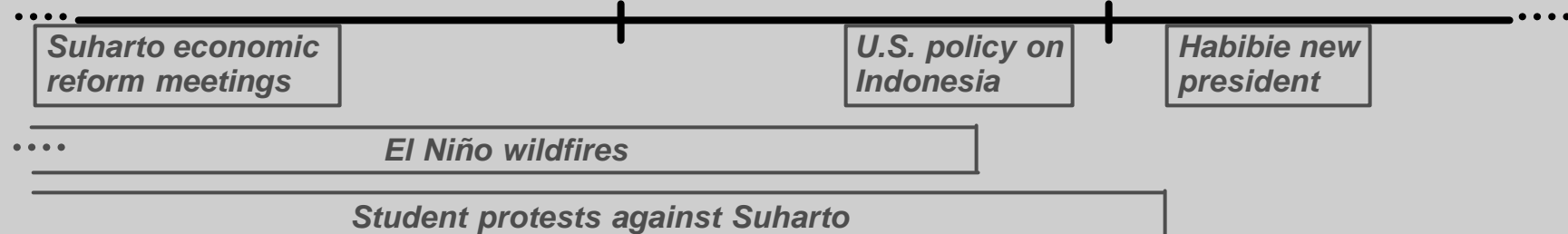
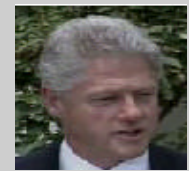
March 1998



April 1998



May 1998



Timeline collage emphasizing “key player faces” and short event descriptors, representing the same data shown in Indonesia map perspective.

# Research Overview

---

- *Collages* as video information synthesis across time, space and sources
- Underlying information extraction and metadata creation
- Adaptive and context-dependent video information presentation
- Multimodal query: Beyond query/browse by text to video exploration
- Interoperable heterogeneous distributed libraries and standards
- External testbeds, corpora and usability evaluations

## Significance of Informedia-II

---

- The use of integrated AI technology for underlying metadata extraction towards video “understanding” continues
- Research is highly integrated and synergistic program
  - experimentation at scale:
    - very large corpus
    - educational users in controlled environments
- Informedia-II begins where Informedia-I ends: with the result set
  - e.g., summarizing, detailing, visualizing, describing in context
- Informedia-II attempts to extend the function of digital libraries
  - combining and distilling information to generate encyclopedia-like comprehensiveness and presentation on-demand

# Related Work and Futures

---

## Related Work

- Long-term goal: Index and search continuously captured, unedited video, therefore ...
- Cognition: object recognition, noisy speech understanding, spoken language translation
- Browsing: topic detection & tracking, text summarization
- Infrastructure: adaptive networks, intelligent storage

## Project Legacy

- No future funding:
  - Industrial or foundation funding of special public domain collections
  - Push back to IP owners of proprietary collections
- With funding:
  - Retention & extension of collections if tied to new research